

МЕТОД ЧАСТОТНО – ВРЕМЕННОГО АНАЛИЗА СИГНАЛОВ

С.П.БЕЛОВ
Е.И. ПРОХОРЕНКО
А.С.БЕЛОВ

*Белгородский
государственный
университет*

e-mail: Belov@bsu.edu.ru

В статье рассматриваются возможности практического применения нового метода частотного анализа, позволяющего вычислять точное распределение энергии по частотным интервалам при любом разбиении частотной области (в том числе и неравномерном) без вычисления трансформант Фурье.

Ключевые слова: частотный анализ, кратковременный спектр, речевой сигнал.

Введение

В настоящее время для анализа речевых сигналов широко используется вычисление мгновенного энергетического спектра, который позволяет получить представление о распределении энергии отрезка речевого сигнала по частотным интервалам, что является важным, например, в задачах распознавания речи.

Существующие методы вычисления частотных характеристик речевых сигналов основываются на использовании традиционных методов дискретного преобразования Фурье, который обладает большой погрешностью [1].

Во-первых, количество комплексных коэффициентов БПФ (результатов преобразования) будет равно количеству обрабатываемых отсчетов сигнала и целесообразно, чтобы оно было равно степени числа два, т.е. 2^n . Поэтому обычно каждый из коэффициентов соотносят с одной из полос равномерного разбиения частотной оси на 2^n интервалов. Иными словами, количество частотных полос и их ширина тесно связаны с длиной обрабатываемого отрезка, что не всегда оправдано.

Во-вторых, квадрат модуля комплексного коэффициента БПФ может служить только оценкой энергии отрезка сигнала, приходящейся на соответствующий интервал. При этом точность такой оценки неизвестна и существенно зависит от соотношений фазовых характеристик составляющих исходного сигнала и дискретизатора.

В статье предлагается метод, позволяющий вычислять точное распределение энергии, причем при любом желаемом разбиении на частотные интервалы (в том числе неравномерном), применение которого дает возможность адекватно учесть свойство сосредоточенности энергии звуков речи в малом количестве достаточно узких частотных интервалах, а также рассматриваются некоторые возможности его практического применения.

Теоретические основы метода

Пусть компоненты вектора $\vec{x} = (x_1, \dots, x_N)^T$ представляют собой значения некоторого сигнала (функции времени), которые соответствуют значениям аргумента $i\Delta t$, т.е.

$$x_i = x(i\Delta t), \quad i = 1, \dots, N,$$

где Δt - интервал дискретизации по времени.

Положим далее

$$X(\nu) = \sum_{k=1}^N x_k e^{-j(k-1)\nu}, \quad (1)$$

т.е. $X(\nu)$ представляет собой трансформанту Фурье отрезка отсчетов сигнала (вектора), в качестве области определения которой рассматривается (нормированная частота)

$$-\pi \leq \nu \leq \pi,$$

так что имеет место обратное преобразование

$$x_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\nu) e^{j(i-1)\nu} d\nu.$$

Отсюда нетрудно получить равенство Парсеваля

$$\|\vec{x}\|^2 = \sum_{k=1}^N x_k^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\nu)|^2 d\nu,$$

так что

$$P_r(\vec{x}) = \frac{1}{2\pi} \int_{\nu \in V_r} |X(\nu)|^2 d\nu \quad (2)$$

представляет собой долю энергии отрезка сигнала (евклидовой нормы вектора), соответствующую частотному интервалу

$$V_r = [-\nu_{2r}, -\nu_{1r}) \cup [\nu_{1r}, \nu_{2r}).$$

Подстановка определения (1) в интеграл (2) позволяет получить выражение вида:

$$P_r(\vec{x}) = \vec{x}^T \mathbf{A}_r \vec{x}, \quad (3)$$

где $\mathbf{A}_r = \{a_{ik}^r\}$, а

$$a_{ik}^r = \begin{cases} \frac{\sin[\nu_{2r}(i-k)] - \sin[\nu_{1r}(i-k)]}{\pi(i-k)}, & i \neq k, \\ \frac{\nu_{2r} - \nu_{1r}}{\pi}, & i = k. \end{cases}$$

Таким образом, долю энергии отрезка сигнала в любом частотном интервале можно вычислить на основе представления (3), не вычисляя при этом, согласно (1), соответствующую трансформанту Фурье.

Очевидно, что соотношения (3) представляет собой новый инструмент, позволяющий анализировать частотно-энергетические характеристики речевых сигналов. При этом выбор ширины и расположения каждого частотного интервала не зависит от длительности окна анализа сигнала.

Из теории матриц известно, что если матрица симметрична, то существует N собственных векторов, которые соответствуют собственным числам λ_k . Матрица \mathbf{A}_r обладает свойством быстрого затухания значений собственных чисел. Вычисления показывают, что при выполнении неравенства $M = 2[N/(2R)] \geq 4$ собственные числа $\lambda_1 \approx \lambda_2 \approx \dots \approx \lambda_M \approx 1$, а $\lambda_{M+k} \approx 0$, $k = 1, 2, \dots, N$. Этим свойством можно воспользоваться для ускорения вычисления частотно-энергетических характеристик речевых сигналов, используя разложение по собственным векторам, собственные числа которых близки к единице [2].

Применение метода при обработке речевых сигналов

1. Обнаружение границы пауза/звук

Задача сегментации речевого сигнала является очень важной в задаче распознавания речи. Она решается как при создании обучающих баз данных, содержащих фразы, так и во время работы систем распознавания слитной речи, основанных на фонемном подходе. Первичная сегментация речевого сигнала (разделение фраз и звукосочетаний) может быть реализована при помощи процедуры обнаружения границы пауза/звук.

Анализ тонкой структуры энергетического спектра речевого сигнала, состоящего из звуков и пауз, показывает, что энергия фрагмента сигнала, соответствующего паузе, распределена относительно равномерно по всему диапазону частот, тогда как энергия фрагмента сигнала, соответствующего звуку, сосредоточена в узких полосах частот. Нами предлагается метод обнаружения пауз, основанный на учете отличий в распределении энергетических составляющих звуков речи и сигнала паузы в частотной области [3].

Формулируется следующая гипотеза:

Но: энергия исходного отрезка \vec{x}_i , $i = 1, \dots, N$ в r -том частотном интервале $(\nu_{2r} - \nu_{1r})$ обусловлена внешними шумами.

Положим

$$S_r^i = \frac{\hat{P}_r^i}{P_r^{\Pi}}, r = 1, 2, \dots, R. \quad (4)$$

В случае, когда i -й анализируемый отрезок сигнала \vec{f}_i представляет собой отрезок паузы, правая часть (4) представляет собой случайную величину приближенно распределенную по закону

$$\chi_K^2: \frac{1}{2^{\frac{M_1}{2}} \Gamma\left(\frac{M_1}{2}\right)} (S_r^i)^{\frac{M_1}{2}-1} e^{-\frac{S_r^i}{2}}, \quad (5)$$

т.е. можно установить порог h_α , который удовлетворяет неравенству

$$\frac{1}{2^{\frac{M_1}{2}} \Gamma\left(\frac{M_1}{2}\right)} \int_{h_\alpha}^{\infty} y^{\frac{M_1}{2}-1} e^{-\frac{y}{2}} dy \leq \alpha, \quad (6)$$

где α - так называемая вероятность ложной тревоги, M_1 - число степеней свободы, которое можно приблизительно вычислить как $M_1 = \sum_{i=1}^M \lambda_i$.

Для увеличения вероятности правильного обнаружения границы пауза/звук целесообразно с порогом сравнивать максимальное значение из отношений вида (4), так как энергия сигнала, соответствующего звуку может быть сосредоточена в сравнительно узком диапазоне частот.

Таким образом, если имеет место

$$\max(S_r^i) \geq h_\alpha, \quad (7)$$

то отвергается нулевая гипотеза H_0 .

Если же выполняется неравенство

$$\max(S_r^i) < h_\alpha, \quad (8)$$

то H_0 считается справедливой и данный отрезок речевого сигнала принимается за паузу.

Для иллюстрации работоспособности метода были проведены вычислительные эксперименты.

При достаточно малой вероятности ложной тревоги и соответствующей величине порога $h_\alpha = 2M_1$, выбранной по таблицам вероятностей распределения χ^2 [4] при $N=60$, $R=10$, число переходов решающей функции $\max(S_r^i)$ через установленный порог на отрезках сигнала соответствующих паузе, не превышает заданную величину. Наиболее типичные результаты приведены в таблицах 1-2. Здесь вероятность ложной тревоги определялась как

$$\alpha = \frac{N_{пер}}{N_{\Pi}},$$

где N_{Π} - длина паузы, $N_{пер}$ - количество переходов решающей функции через порог.

Таблица 1

Длина паузы, M_{Π}	30150	25150	14150	6150	14150	20150
Количество переходов решающей функции через порог, $M_{пер}$	1560	1480	0	130	380	590
Вероятность ложной тревоги, α	0,015	0,019	0	0,02	0,018	0,02

С другой стороны, на участках соответствующих звукам речи решающая функция превышала установленный порог с вероятностью 90-95%. В таблице 2 приведены результаты исследования поведения решающей функции на отрезках сигнала, соответствующих слитной речи.

Таблица 2

Длина фрагмента слитной речи, M_p	4614	17863	8573
Количество переходов решающей функции через порог, $M_{пер}$	4256	17002	7741
Вероятность правильного обнаружения, $P_{по}$	0.96	0.97	0.96

Причем, отрезки сигнала, на которых величина решающей функции не превышала установленный порог, как правило, являлись паузами малой длительности между фонемами, или слитно произнесенными словами, например, как показано на рис. 1. Здесь приведена фраза «правила пользования Интернет» произнесенная слитно, и отрезок, определенный как пауза, соответствует паузе между фрагментами фразы «правила», «пользования» и «интернет».

На рис. 2 показана граница пауза/звук определенная в начале фразы.

В другом случае (рис. 3) анализировалась фраза, содержащая слово «пытаются». Здесь отрезки, определенные как паузы, соответствуют паузам между звуко сочетаниями «пы» - «тают» - «ся».

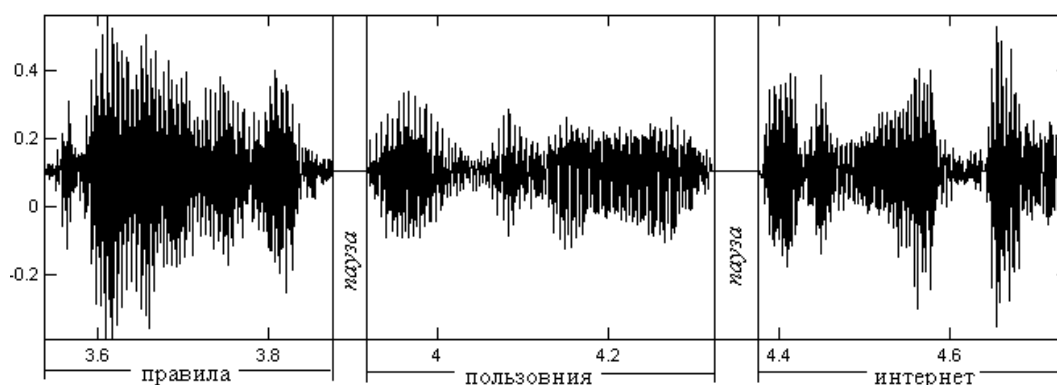


Рис. 1. Обнаружение паузы между словами фразы

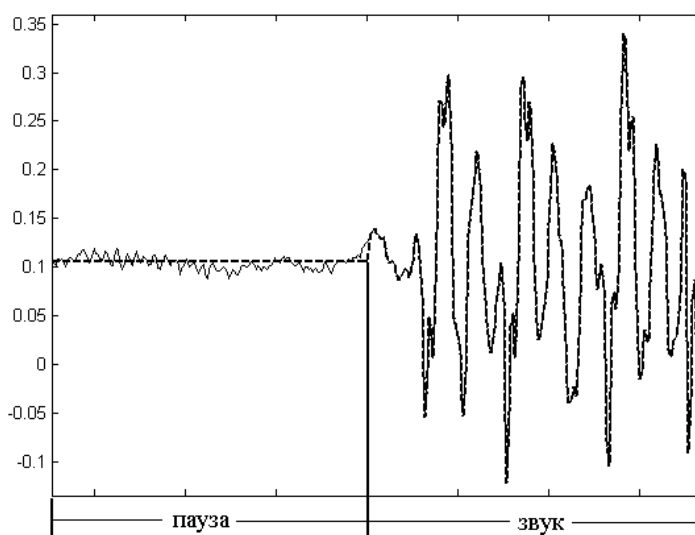


Рис. 2. Граница пауза/звук, определенная в начале фразы

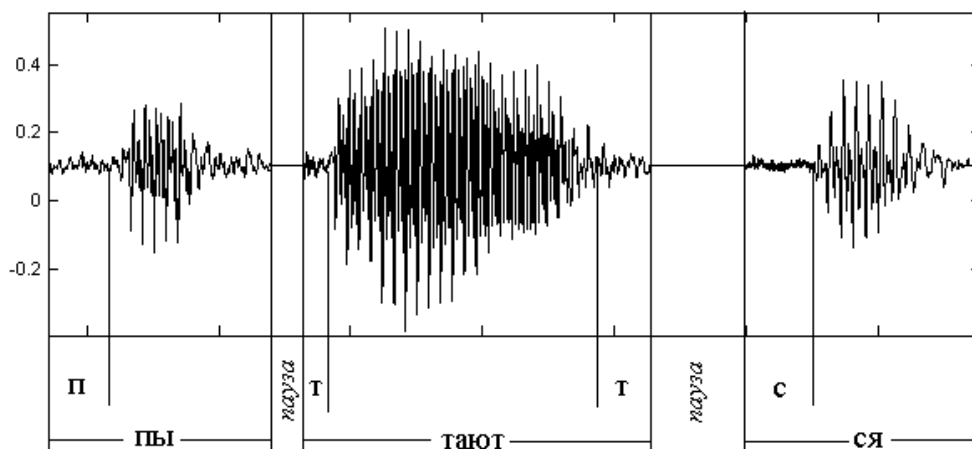


Рис. 3. Обнаружение пауз между звуко сочетаниями

В состав этого слова входят звуки «п», «т», которые принадлежат к невокализованным звукам малой длительности, вероятность пропуска которых наиболее велика, из-за их малого (по сравнению с вокализованными звуками) уровня, а так же «с» - шумоподобный звук, и в данном случае отрезки, определенные как паузы, приходятся на начало и окончание звука «т» и начало звука «с». Срезание начала звуков в этих случаях особенно нежелательно, так как это может снизить разборчивость речи. Визуально сложно определить, на какой части этих звуков значение решающей функции не превышало установленного порога, но экспертная оценка при воспроизведении сигнала с удаленными паузами показала, что звуки «п», «т» и «с» в слове «пытаются» четко различимы.

2. Вычисление спектрограмм

Кроме одномерных спектров (амплитуда-частота) широко используются, в том числе в задачах распознавания речи, динамические спектральные развертки речевого сигнала, называемые спектрограммами, то есть частотно-временные описания мгновенных спектров речи с заданным шагом наблюдения (анализа) по времени и по частоте. Спектрограммы позволяют получить значительно больше информации не только о спектральном составе речевых сигналов, но и о характере их изменения во времени.

Существующие методы вычисления спектрограмм также основываются на традиционных методах дискретного преобразования Фурье, о некоторых недостатках которого говорилось выше. Соотношение (3) представляет собой новый инструмент, позволяющий реализовать новую процедуру вычисления спектрограмм, адекватно учитывающую свойство сосредоточенности энергии звуков речи в малом количестве достаточно узких частотных интервалах.

Результаты проведенных экспериментов позволяют утверждать, что при увеличении количества частотных интервалов погрешность вычисления долей энергии речевого сигнала с использованием ДПФ, по сравнению с предлагаемым методом возрастает и, как следствие, ухудшается точность представления звуков при построении спектрограмм [5].

Далее приведены спектрограммы, построенные для фрагментов речевых сигналов различной длительности, с использованием соотношения (3).

На рис. 4,а представлен фрагмент речевого сигнала соответствующий слову «аппроксимация» длительностью 1,49 с и его спектрограмма (рис. 4,б) при разбиении диапазона на 128 равных частотных интервалов и длине окна анализа 512 отсчетов. В данном случае узкополосная спектрограмма (частота развертки 30Гц) позволяет отчетливо увидеть гармоники голосового источника и формантные частоты звуко сочетаний. На рис. 4,в представлен фрагмент речевого сигнала соответствующий звуку «а» длительностью 214 мс и его спектрограмма при разбиении диапазона на 128 частотных интервалов (рис. 4,г) и длине окна анализа 512 отсчетов. На спектрограмме отчетливо видны формантные частоты звука в период времени с 0 до 120 мс и изменение частотно-энергетических характеристик звука в период времени с 120 до 214 мс, что соответствует изменению формы самого сигнала (окончание звука «а»).

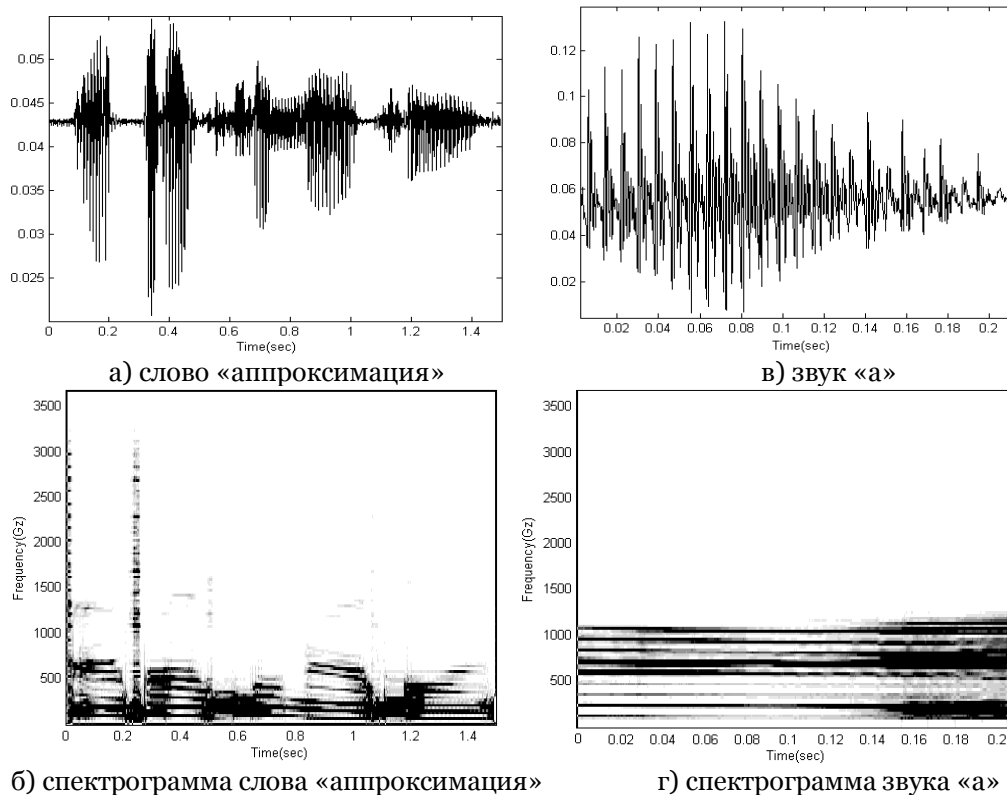


Рис. 4. Спектрограммы речевых сигналов ($N=512$, $R=128$)

Таким образом, новый метод вычисления спектрограмм позволяет определять точные значения доли энергии отрезка сигнала, попадающего в любой конечный частотный интервал, и, следовательно, обладает значительно большей эффективностью по сравнению с вычислением спектрограмм традиционным методом дискретного преобразования Фурье.

Литература

1. Жилияков, Е.Г. О сжатии речевых сигналов. [Текст] / Е.Г. Жилияков, С.П. Белов, Е.И. Прохоренко // Вестник Национального технического университета «ХПИ». – Харьков, 2005. – вып.56. – С. 32-41. – (Сборник научных трудов. Тематический выпуск: Информатика и моделирование).
2. Жилияков, Е.Г. Вариационные методы частотного анализа звуковых сигналов [Текст] / Е.Г. Жилияков, С.П. Белов, Е.И. Прохоренко // Труды учебных заведений связи. – СПб, 2006. – № 174. – С.163-170.
3. Жилияков, Е.Г. Новый метод сжатия речевых данных [Текст] / Е.Г. Жилияков, С.П. Белов, Е.И. Прохоренко // Труды учебных заведения связи. – СПб, 2006. – №175. – С. 152-161.
4. Таблицы математической статистики [Текст] / Л.Н. Большев, Н.В. Смирнов. – М.: Наука. Главная редакция физико-математической литературы, 1983. – 416с.
5. Жилияков, Е.Г. Новый метод вычисления спектрограмм речевых сигналов [Текст] / Е.Г. Жилияков, С.П. Белов, Е.И. Прохоренко // Материалы XIII международной научно-технической конференции «Радиолокация, навигация, связь». Воронеж, 2007. – С. 278-283.

METHOD FREQUENCY – TEMPORAL ANALYSIS OF SIGNALS

S.P.BELOV
E.I. PROHORENKO
A.S.BELOV

Belgorod State University

e-mail: Belov@bsu.edu.ru

Possibilities of practical application of new method of frequency analysis, allowing to calculate the exact distributing of energy on frequency intervals at any breaking up of frequency area (including uneven) without the calculation of transforms Fur'e are examined in the article.

Keywords: frequency analysis, brief spectrum, vocal signal.